Mol Gen Genet (1987) 210:437–442

MGG

Cloning of an Arabidopsis thaliana gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants

Harry J. Klee, Yvonne M. Muskopf, and Charles S. Gasser
Plant Molecular Biology Group, Monsanto Company, 700 Chesterfield Village Parkway, Chesterfield MO 63198, USA

Summary. 5-enolpyruvylshikimate-3-phosphate synthase (EPSPs), the target of the herbicide glyphosate, catalyzes an essential step in the shikimate pathway common to aromatic amino acid biosynthesis. We have cloned an EPSP synthase gene from *Arabidopsis thaliana* by hybridization with a petunia cDNA probe. The *Arabidopsis* gene is highly homologous to the petunia gene within the mature enzyme but is only 23% homologous in the chloroplast transit peptide portion. The *Arabidopsis* gene contains seven introns in exactly the same positions as those in the petunia gene. The introns are, however, significantly smaller in the *Arabidopsis* gene. This reduction accounts for the significantly smaller size of the gene as compared to the petunia gene. We have fused the gene to the cauliflower mosaic virus 35 *S* promoter and reintroduced the chimeric gene into *Arabidopsis*. The resultant overproduction of EPSPs leads to glyphosate tolerance in transformed callus and plants.

Key words: Shikimate pathway – Herbicide – Transgenic plants – EPSP synthase

Introduction

The enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPs) is a critical enzyme in aromatic amino acid biosynthesis, catalyzing the addition of the enolpyruvyl moiety of phosphoenolpyruvate to shikimate-3-phosphate. This enzyme is the target of the potent, broad-spectrum herbicide, glyphosate [N-(phosphonomethyl)glycine] (Steinrücken and Amrhein 1980). Inhibition of EPSPs by glyphosate prevents the synthesis of chorismate-derived aromatic amino acids and secondary metabolites. Although the EPSPs gene is nuclear, the mature enzyme is localized in the chloroplast (Mousdale and Coggins 1985; della-Cioppa et al. 1986). EPSPs is synthesized as a preprotein, containing a transit peptide, by free cytoplasmic ribosomes. The precursor is transported into the chloroplast stroma and proteolytically processed to yield the mature enzyme (della-Cioppa et al. 1986).

Recently, a gene for petunia EPSPs has been cloned and used to engineer glyphosate tolerance in transgenic petunia plants (Shah et al. 1986). This was accomplished by fusing a cDNA clone to the cauliflower mosaic virus (CaMV) 35 *S* promoter, resulting in several-fold overproduction of the enzyme. Because of the importance of this enzyme in developing herbicide tolerance, there has been a great deal of interest in understanding the enzymology of EPSPs and its inhibition by glyphosate. One approach to understanding the important components of an enzyme is to determine the primary sequences from several different species and to compare the conserved and divergent amino acids. This approach, in the case of EPSPs, simultaneously provides useful information concerning the requirements for sequestration of nuclear-encoded proteins in the chloroplast.

Arabidopsis thaliana, a member of the *Brassicaceae*, represents a unique tool to the plant molecular biologist. The properties which make this organism so amenable to molecular genetic manipulation have been thoroughly reviewed (Estelle and Somerville 1986; Meyerowitz and Pruitt 1986). Briefly, this cruciferous weed has a very small genome of only about 7×10^7 bp (Leutwiler et al. 1984). It contains very little repetitive DNA in comparison with other plants (Pruitt and Meyerowitz 1986). There is a detailed genetic map available with numerous, well characterized mutations (Koornneef et al. 1983). Recently, an *Agrobacterium*-mediated transformation system which allows for introduction of foreign genes has been developed (Lloyd et al. 1986).

The small genome size and lack of repetitive DNA make *Arabidopsis* an excellent system for the isolation of genes. Screening of a small number of genomic clones is sufficient to isolate a gene of interest. The small genome size may result in smaller genes and fewer members of multigene families. In at least one case this seems to be true. While the extensin gene family may have up to 500 members in petunia, there appear to be only 5–6 genes in *Arabidopsis* (D. Shah, personal communication). We report here on the cloning and gene organization of an *Arabidopsis* EPSPs. This gene is, in fact, considerably smaller than its counterpart in petunia. The ease of cloning and characterization of this gene illustrate the major advantage of the *Arabidopsis* system. The *Arabidopsis* EPSPs gene has been fused to the CaMV 35 *S* promoter and reintroduced into *Arabidopsis* by *Agrobacterium*-mediated transformation. The resultant overproduction leads to glyphosate-tolerant transformed tissue and plants.

Materials and methods

Genomic bank construction. A genomic bank of *Arabidopsis thaliana* (L.) Heynh. (kindly provided by C. Somerville) was constructed in the bacteriophage lambda vector

Offprint requests to: C.S. Gasser

EMBL3 (Frischauf et al. 1983). Genomic DNA was prepared according to Kislev and Rubenstein (1980) and partially cleaved with MboI. The cut DNA was size fractionated by agarose gel electrophoresis and the fragments between 16–20 kb were recovered by electrophoresis onto NA-45 membrane (Schleicher and Schuell). The purified DNA was ligated to EMBL3 arms and packaged in vitro with commercially available packaging extract (Stratagene). Packaged DNA was introduced into *Escherichia coli* C600 (F^- *thi-1 thr-1 leuB6 lacY1 tonA21 supE44* λ^-).

Identification of an EPSPs genomic clone. Approximately ten thousand phage were plated on a 150 mm petri plate and the resultant plaques were lifted onto nitrocellulose filters (Schleicher and Schuell). Filters were hybridized with a nick-translated petunia EPSP synthase cDNA probe which had been isolated from pMON9566 (C. Gasser, unpublished). This probe contains the entire coding sequence of the mature petunia protein. Hybridization was done in 6 × SSC (1 × SSC is 0.15 M NaCl, 15 mM sodium citrate), 50 mM Tris, pH 7.0, 10 mM EDTA, 0.1% SDS and 100 µg/ml denatured salmon sperm DNA at 65° C for 16 h. Filters were washed 4 × 30 min in 2 × SSC, 0.1% SDS and once for 15 min in 0.2 × SSC, 0.1% SDS. Two plaques which hybridized under these conditions were purified to homogeneity.

Two *Bgl*II fragments of 6.0 and 3.2 kb which hybridized to the petunia gene probe were subcloned from one of the phage into pMON550. Plasmid pMON550 is a derivative of pUC19 (Yanisch-Perron et al. 1985) which contains a unique *Bgl*II site in the polylinker (H. Klee, unpublished). Smaller fragments were then subcloned into either pUC118 or pUC119 (kindly provided by J. Vieira) for DNA sequence determination. Sequences of the subclones were determined using the dideoxy procedure of Sanger et al. (1977). Sequence comparisons were performed with the University of Wisconsin Genetics Computer Group programs. The genomic EPSPs hybridization pattern was determined by restriction endonuclease digestion of DNA, electrophoresis on a 0.8% agarose gel and Southern blotting (Southern 1975). Blotted DNA was hybridized with nick-translated probe for 16 h in 6 × SSC and washed four times for 15 min in 2 × SSC at 68° C.

Reintroduction and overexpression of the EPSPs gene in Arabidopsis. The *Arabidopsis* EPSPs gene was subcloned into a vector suitable for introduction into and overexpression in plants as follows: the 5′ end of the gene from nucleotides 314 to 693 (Fig. 1) was cloned as a *Sna*BI to *Bam*HI fragment into bluescript KS^+ (Stratagene) creating pMON9734. The 3′ end of the gene (nucleotides 1423 to 4160, Fig. 1) was then cloned into pMON9734 as a *Bam*HI to *Eco*RI fragment, creating pMON588. The central 800 bp *Bam*HI fragment was then inserted into pMON588, reconstructing the intact coding sequence as pMON597. The coding sequence was excised from this plasmid as a *Cla*I to *Eco*RI fragment and ligated into the plant transformation vector pMON857 (M. Hayford and H. Klee, unpublished). This vector contains a CaMV 35 *S* promoter/nopaline synthase 3′ cassette for expression of foreign coding sequences

```
ccctt catgtcttttgtagaaacccattatctttcttagggcccaattgaaaacccaca   60
ttttctttcacctaacccaccaaagccttgcacatgttgacgtgaacaccaaactaacac  120
gtgtcatactgccagtggttatgataaatgctcataccataccagagtcatagagttttt  180
ggttggtgaaagatttgacggatgccttcttctcatttctcaccaactccctccaaaccc  240
aacaaaatgtttatattagcaaagccgccaaagtgtaaacgaaagtttataaatttcatt  300
tctgtgatcttacgtaattggaggaagatcaaaattttcaatccccattcttcgattgct  360
tcaattgaagtttctccgATGGCGCAAGTTAGCAGAATCTGCAATGGTGTGCAGAACCCA  420
                    MetAlaGlnValSerArgIleCysAsnGlyValGlnAsnPro

TCTCTTATCTCCAATCTCTCGAAATCCAGTCAACGCAAATCTCCCTTATCGGTTTCTCTG  480
SerLeuIleSerAsnLeuSerLysSerSerGlnArgLysSerProLeuSerValSerLeu

AAGACGCAGCAGCATCCACGAGCTTATCCGATTTCGTCGTCGTGGGGATTGAAGAAGAGT  540
LysThrGlnGlnHisProArgAlaTyrProIleSerSerSerTrpGlyLeuLysLysSer

GGGATGACGTTAATTGGCTCTGAGCTTCGTCCTCTTAAGGTCATGTCTTCTGTTTCCACG  600
GlyMetThrLeuIleGlySerGluLeuArgProLeuLysValMetSerSerValSerThr

GCGGAGAAAGCGTCGGAGATTGTACTTCAACCCATTAGAGAAATCTCCGGTCTTATTAAG  660
AlaGluLysAlaSerGluIleValLeuGlnProIleArgGluIleSerGlyLeuIleLys

CTTCCTGGCTCCAAGTCTCTATCAAATCGGATCCTGCTTCTCGCTGCTCTGTCTGAGgta  720
LeuProGlySerLysSerLeuSerAsnArgIleLeuLeuLeuAlaAlaLeuSerGlu

tatatcacttcgtttcgtccttctctgtaatctgaacttagattataaagattgatactt  780

taccattttgctgtggttttatagGGAACAACTGTAGTGGACAACTTGTTGAATAGCGAT  840
                          GlyThrThrValValAspAsnLeuLeuAsnSerAsp

GACATCAATTACATGCTTGATGCGTTGAAGAGATTGGGACTTAATGTGGAAACTGACAGT  900
AspIleAsnTyrMetLeuAspAlaLeuLysArgLeuGlyLeuAsnValGluThrAspSer

GAAAATAATCGTGCTGTAGTTGAAGGATGTGGCGGGATATTCCCAGCTTCCATAGATTCA  960
GluAsnAsnArgAlaValValGluGlyCysGlyGlyIlePheProAlaSerIleAspSer

AAGAGTGATATCGAACTTTACCTCGGTAATGCAGGAACAGCAATGCGTCCACTTACCGCT 1020
LysSerAspIleGluLeuTyrLeuGlyAsnAlaGlyThrAlaMetArgProLeuThrAla

GCGGTCACTGCTGCAGGTGGAAACGCAAGgtagattgaaggagttgatgcttcttggtat 1080
AlaValThrAlaAlaGlyGlyAsnAlaArg

ttgatgtttaaggaatggagcttttgttgatgctttatgatccatttatttccagTTATGT 1140
                                                      TyrVal

GCTTGATGGGGTGCCTCGTATGAGAGAAAGACCTATAGGGGATTTGGTTGTTGGTCTTAA 1200
 LeuAspGlyValProArgMetArgGluArgProIleGlyAspLeuValValGlyLeuLys

GCAGCTTGGTGCTGATGTTGAATGTACTCTTGGAACTAACTGCCCTCCTGTTCGTGTCAA 1260
 GlnLeuGlyAlaAspValGluCysThrLeuGlyThrAsnCysProProValArgValAsn

CGCTAATGGTGGCCTTCCCGGTGGAAAGgttagatcttgcaaatggcatgtgaatatgta 1320
 AlaAsnGlyGlyLeuProGlyGlyLys

atctcgttcctactctatgaacacttgcagaaatgtgtgttcatcatagcctagcttg  1380
acaagatttcagtttttaatctactctcaacggatggatcctaaaatagaatcggatttg  1440
gtgattggttttcgttctcgattaccgttttcgttgtatgatttcttgattaacaattag  1500
gagacatgttatgcatttgcagGTGAAGCTTTCTGGATCAATTAGTAGTCAGTACTTGAC  1560
                     ValLysLeuSerGlySerIleSerSerGlnTyrLeuThr

TGCTCTGCTCATGTCTGCTCCCTTAGCTCTTGGAGACGTCGAGATTGAGATTGTCGATAA 1620
 AlaLeuLeuMetSerAlaProLeuAlaLeuGlyAspValGluIleGluIleValAspLys

ATTAATTTCTGTTCCATATGTTGAAATGACATTGAAGTTGATGGAACGTTTCGGGGTTAG 1680
 LeuIleSerValProTyrValGluMetThrLeuLysLeuMetGluArgPheGlyValSer

TGTCGAGCATAGTGATAGCTGGGATCGTTTCTTTGTCAAGGGCGGGCAAAAATACAAgta 1740
 ValGluHisSerAspSerTrpAspArgPhePheValLysGlyGlyGlnLysTyrLys

ggagttattcttttcttcctttttctgaaatcacatccctagcttgacaatataatgact  1800
aaaaggtgaatgattcagGTCTCCGGGTAATGCGTATGTAGAAGGTGATGCTTCTAGTGC  1860
                    SerProGlyAsnAlaTyrValGluGlyAspAlaSerSerAla

TAGTTATTTCTTGGCTGGTGCTGCCATTACCGGTGAAACTGTCACAGTCGAAGGTTGTGG 1920
 SerTyrPheLeuAlaGlyAlaAlaIleThrGlyGluThrValThrValGluGlyCysGly

AACTACCAGCTTGCAGgtaatatttgtacactgaatcatcgacgaggctgttaagtttat 1980
 ThrThrSerLeuGln

agtgaaattcgtctaggtcaaagtttcatcttttgacaagttgtatataacatattcgca 2040
agatctaagctcaatttttgtgatgaatctctagGGAGATGTAAAATTCGCCGAGGTCCT 2100
                              GlyAspValLysPheAlaGluValLeu

TGAGAAAATGGGATGTAAAGTGTCCTGGACAGAGAACAGTGTGACTGTGACAGGACCACC 2160
 GluLysMetGlyCysLysValSerTrpThrGluAsnSerValThrValThrGlyProPro

TAGAGATGCTTTTGGAATGAGACACTTGCGGGCTATTGATGTCAACATGAACAAAATGCC 2220
 ArgAspAlaPheGlyMetArgHisLeuArgAlaIleAspValAsnMetAsnLysMetPro

TGATGTAGCCATGACCCTTGCCGTCGTTGCTCTCTTTGCTGACGGTCCAACCACCATTAG 2280
 AspValAlaMetThrLeuAlaValValAlaLeuPheAlaAspGlyProThrThrIleArg

AGATGgtaagtaaaaagctctctcttataattaaggtttctcaatattcatgatcactta 2340
 Asp

attctgtttggttaatatagTGGCTAGCTGGAGAGTAAAGGAGACAGAAAGGATGATTGC 2400
                     ValAlaSerTrpArgValLysGluThrGluArgMetIleAla

CATTTGCACAGAGCTTAGAAAAgtaagagattcttatctctctttctgtctcttgaca 2460
  IleCysThrGluLeuArgLys

gtgctcattctaagtaattagctcataaatttgtgtgtttgtgttcagCTGGGAGCTACA 2520
                                          LeuGlyAlaThr

GTGGAAGAAGGTTCAGATTATTGTGTGATAACTCCGCCCAAAAAGGTGAAAACGGCAGAG 2580
ValGluGluGlySerAspTyrCysValIleThrProProLysLysValLysThrAlaGlu

ATTGATACATATGATGATCATAGAATGGCAATGGCATTCTCTCTTGCAGCTTGTGCTGAT 2640
IleAspThrTyrAspAspHisArgMetAlaMetAlaPheSerLeuAlaAlaCysAlaAsp

GTTCCAATCACCATCAACGACTCTGGTTGCACCAGGAAAACCTTCCCCGACTACTTCCAA 2700
ValProIleThrIleAsnAspSerGlyCysThrArgLysThrProAspTyrPheGln

GTACTTGAAAGAATCACAAAGCACTAAacaataaactctgtttttttcttctgatccaagc 2760
ValLeuGluArgIleThrLysHisEnd

tt  2762
```

Fig. 1. Nucleotide sequence of the *Arabidopsis* EPSPs gene. The deduced exon sequences are shown in *upper case letters*. The deduced amino acid sequence is indicated below the DNA sequence

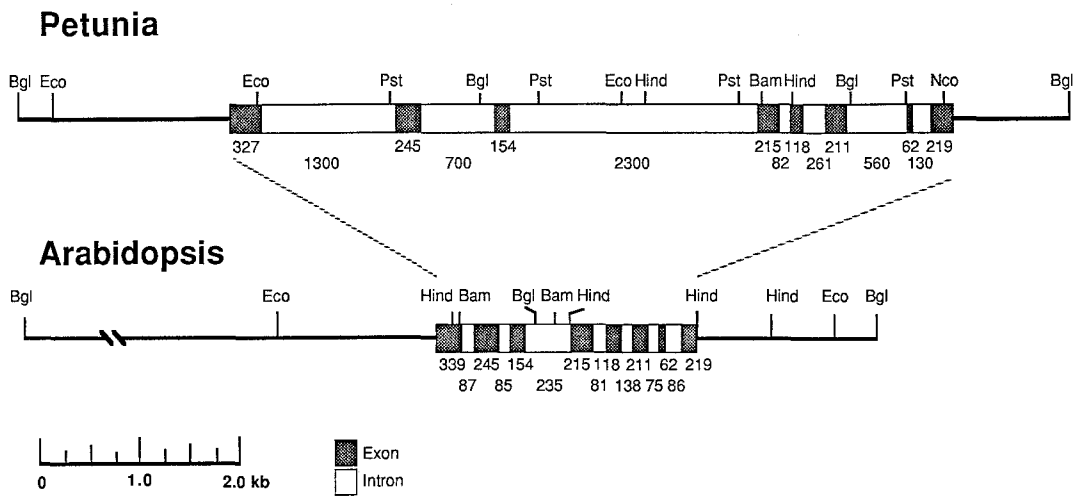

Fig. 2. The genomic organization of the *Arabidopsis* EPSP synthase gene compared to that of petunia. *Numbers* refer to the sizes (in nucleotides) of the exons and introns

as well as a selectable marker encoding resistance to gentamicin in plants. The resulting plasmid, pMON599, was introduced into a disarmed *Agrobacterium* helper strain containing pTiT37SE (Rogers et al. 1987) and used to transform *Arabidopsis* by the method of Lloyd et al. (1986). Selection for transformed tissue was accomplished by growing callus om MS104 plates (Ausubel et al. 1980) containing 40 μg/ml gentamicin (Sigma). Details of the construction and use of gentamicin resistance vectors in plant transformation will be provided elsewhere (M. Hayford and H. Klee, in preparation). Gentamicin-resistant transformed callus was tested for glyphosate tolerance by plating small pieces (approximately 5 mm in diameter) on MS104 plates containing from 0 to 2.0 mM glyphosate. Progeny of transformed, regenerated plants were tested for inheritance of the EPSPs gene by germinating sterile seeds on plates containing half-strength MS salts and 0.5 mM glyphosate.

Results

We were interested in cloning the EPSPs gene from *Arabidopsis* for several reasons. First, we wanted to obtain an EPSPs gene from an organism other than petunia to demonstrate that overexpression could be used as a general mechanism for engineering glyphosate tolerance. Second, we were interested in comparing the amino acid sequences from members of two divergent plant families to determine which sequences of the transit peptides and mature enzymes are important for function. Finally, we were interested in examining the structure of an *Arabidopsis* gene to determine whether the smaller genome size would result in a gene with smaller or fewer introns than a plant with a much larger genome size such as petunia.

To isolate the *Arabidopsis* EPSP synthase gene, we constructed a genomic library and screened approximately ten thousand plaques with a petunia EPSPs probe. Of the five plaques which hybridized to the petunia probe, two were purified to homogeneity and were found to be overlapping clones of the same gene.

Restriction endonuclease fragments from the purified phage which hybridized to the petunia gene probe were subcloned and further characterized by restriction and DNA sequence analysis. The DNA sequence of the *Arabidopsis* gene is shown in Fig. 1. The organization of the cloned *Arabidopsis* gene as deduced from the sequence analysis is shown in Fig. 2. The intron/exon organization of the *Arabidopsis* EPSPs was determined by comparison of the DNA sequence with those of the petunia gene and cDNA (Gasser et al., in press). The extensive homology within the coding sequences of the two genes allowed for unambiguous assignment of intron/exon boundaries within the *Arabidopsis* gene. These assignments conform in every case to the GT/AG rule (Brown et al. 1986).

In order to map the start point(s) of transcription of the gene, an oligonucleotide complementary to nucleotides 417 to 432 (see Fig. 1) was synthesized. This probe was used to perform a primer extension experiment of poly(A)$^+$ RNA isolated from wild-type plants. Results indicated that transcription of the gene is initiated at multiple sites upstream from the translational start site with no single favored site (data not shown). This lack of a discrete start site is consistent with the results obtained in petunia plants where several start sites were observed (Gasser et al., in press).

To test for the presence of additional EPSPs genes in *Arabidopsis*, a *Bam*HI fragment, internal to the cloned gene (nucleotides 693 to 1421, Fig. 1), was nick-translated and hybridized to a genomic Southern blot of wild-type DNA which had been cut with either *Bam*HI, *Eco*RI or *Hin*dIII. The result of this hybridization is shown in Fig. 3. Fragments corresponding to the cloned gene are labelled with arrows. This blot indicates the presence of additional hybridizing sequences in the *Arabidopsis* genome. At this point it is not known whether the other hybridizing fragments correspond to a functional gene.

Overproduction of wild-type petunia EPSP synthase has previously been demonstrated to confer glyphosate tolerance in petunia plants (Shah et al. 1986). To demonstrate that overexpression could be used as a general method for conferring glyphosate tolerance in other plants, the *Arabidopsis* EPSPs gene was cloned behind the CaMV 35 *S* promoter in a plant transformation vector which would allow for its reintroduction into *Arabidopsis* by means of *Agrobacterium*-mediated transformation. This transformation vector, pMON599 (Fig. 4), contains the chimeric 35 *S*/EPSPs gene as well as a gene conferring gentamicin resistance in

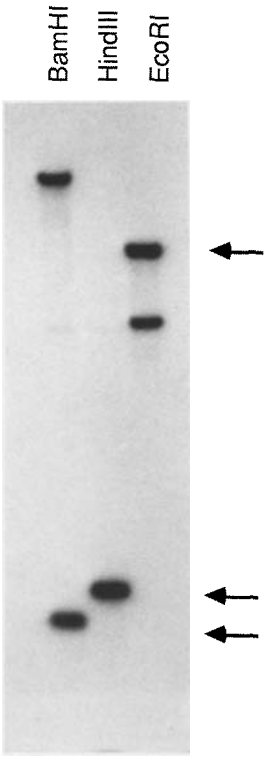

Fig. 3. Southern blot of wild-type *Arabidopsis* DNA digested with *Eco*RI, *Hin*dIII, or *Bam*HI and hybridized to a *Bam*HI fragment encompassing nucleotides 693 to 1421 of the *Arabidopsis* gene. Bands corresponding to the restriction map of the sequenced gene which were expected to hybridize to the *Bam*HI fragment probe are labelled with arrows

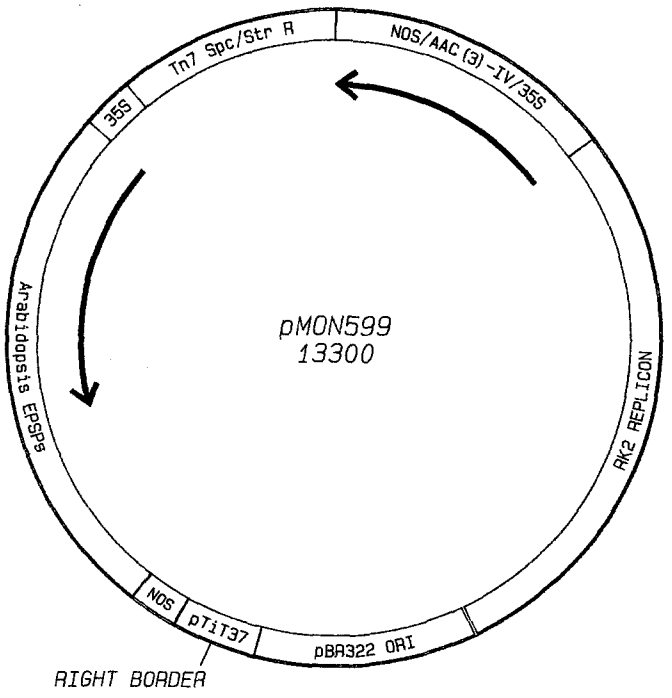

Fig. 4. Structure of pMON599. The plasmid is an *Agrobacterium* binary-type plant transformation vector encoding gentamicin resistance [AAC(3)-IV] (M. Hayford and H. Klee, unpublished). It contains plasmid replication origins for use in *Escherichia coli* (pBR322) and *Agrobacterium* (RK2 REPLICON), a gene encoding bacterial spectinomycin resistance (Tn7 *Spc/Str R*), a T-DNA border sequence derived from pTiT37, and an expression cassette

transformed plant tissue. Transformed gentamicin-resistant callus was tested for its ability to grow on increasing levels of glyphosate, as compared to control tissue which had been transformed with a gentamicin resistance vector lacking the EPSP synthase gene. Fig. 5a shows the result of the assay. While tissue transformed with both plasmids was able to grow well on medium containing 40 μg/ml of gentamicin, only the pMON599-containing tissue grew on medium containing 2.0 mM glyphosate. Seeds from transformed, regenerated plants were also tested for their ability to germinate and grow on glyphosate-containing medium. The results of the germination assay are shown in Fig. 5b. The plant from which the seed was obtained was heterozygous for the T-DNA. Therefore, segregation of the resistant phenotype can be observed. These results indicate that overexpression of the *Arabidopsis* gene did confer the capacity to grow on an otherwise lethal dose of the herbicide. The results demonstrate that overexpression of EPSP synthase as a means of engineering glyphosate tolerance can work with plants and genes which are unrelated to the *Solanaceae*.

Discussion

The results presented here indicate that we have cloned a functional EPSPs gene from *Arabidopsis* by homology to a petunia gene probe. The proteins encoded from these two genes are highly conserved. A comparison of the deduced amino acid sequences of the *Arabidopsis* and petunia genes reveals extensive homology over the length of the entire mature protein (Fig. 6). Within the mature protein (amino acids 77–520) there is 84% identity between the two genes. Many of the changes between the two are also conservative substitutions. In contrast, the chloroplast transit peptides are very divergent at the amino acid level, sharing only 23% homology. The exact site of transit peptide cleavage has not been determined for the *Arabidopsis* protein and is deduced from a comparison to the known cleavage site in petunia. There does appear to be amino acid conservation around the transit peptide cleavage site. Although the two transit peptides are very divergent they are functionally equivalent since we have been able to demonstrate heterologous expression of both of these enzymes in the reciprocal hosts (data not shown). This suggests that, except for the recognition site of the transit peptidase, the secondary structure of a transit peptide is more critical for proper function than the primary amino acid sequence. This observation is consistent with the recent results reported by Baker and Schatz (1987) which indicate that many sequences are capable of targetting a protein to mitochondria. Analysis of additional chloroplast transit peptides will better clarify the nature of the important structures for this function.

It is particularly interesting to compare the organizations of the petunia and *Arabidopsis* EPSPs genes in light of what is known about the genomes of these two organisms. While *Arabidopsis* has a relatively small genome size (approximately 7×10^7 bp) and little highly repetitive DNA, petunia has a very large genome (approximately 6×10^8 bp) (Bennett et al. 1982) and a great deal of repetitive DNA. Results from Southern blotting experiments are consistent with two EPSPs genes in *Arabidopsis*, although we do not know whether the second homologous sequence is a functional gene. Southern analysis of petunia genomic DNA

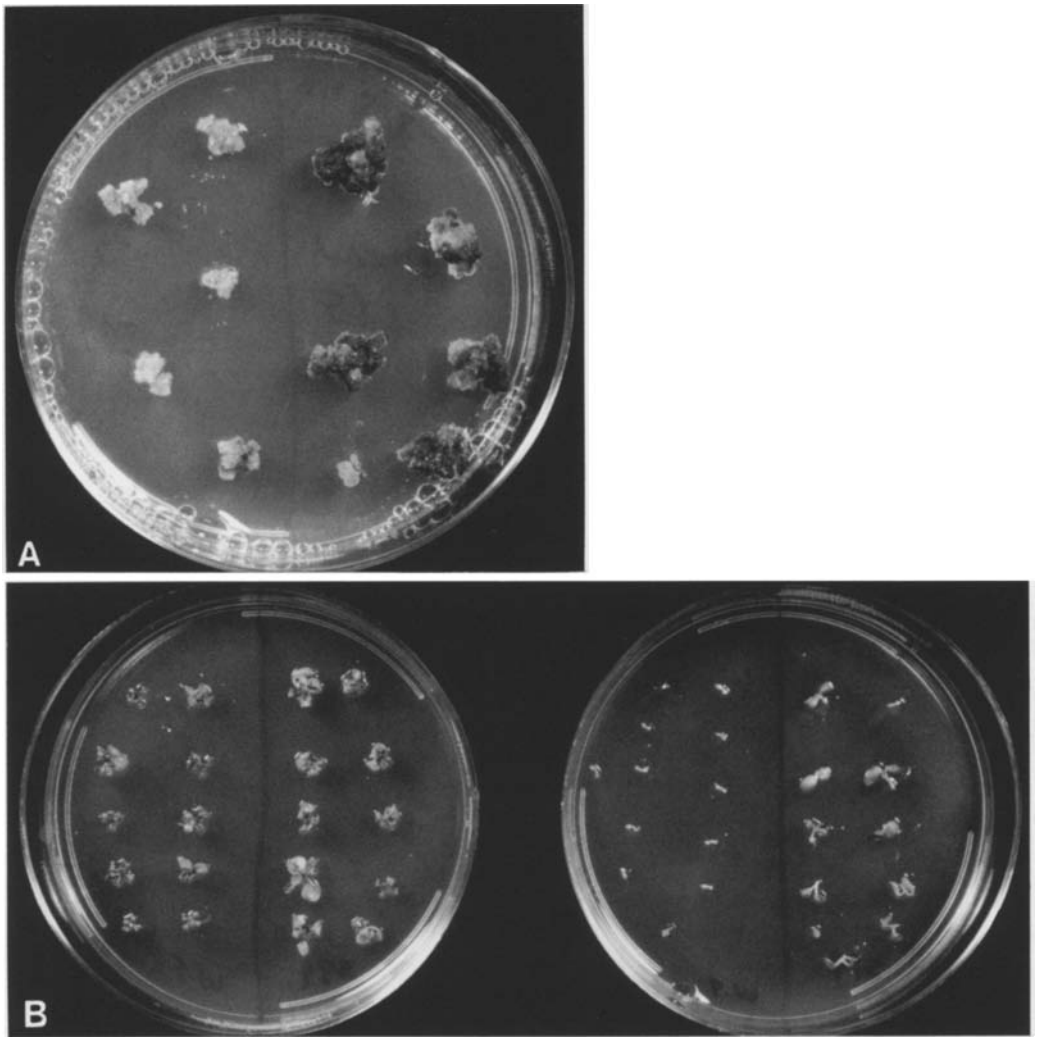

Fig. 5. A Callus assay of transformed *Arabidopsis* tissue for glyphosate tolerance. Transformed callus (*right half of plate*) or callus transformed with a vector not containing EPSPs (*left half of plate*) was placed on medium containing 2.0 mM glyphosate and allowed to grow for 30 days. **B** Seed germination assay. Seeds from self-pollinated transformed plants (*right half of each plate*) or plants transformed with a vector not containing EPSPs (*left half of each plate*) were surface sterilized and germinated on medium containing 0 (*left*) or 0.5 mM glyphosate. The photograph was taken 14 days after germination

indicates two sets of hybridizing sequences as well. The characterized EPSPs genes in both petunia (Gasser et al., in press) and *Arabidopsis* contain seven introns. These introns are in exactly the same positions in both genes (Fig. 3). However, introns in *Arabidopsis* are in every case significantly smaller than in petunia. For example, intron 3 in petunia is 2,300 bp while it is only 150 bp in *Arabidopsis*. Thus, the coding region of the EPSPs, including introns, has been reduced from 7.4 kb in petunia to 2.5 kb in *Arabidopsis*. An examination of different genes from these organisms may provide clues concerning the nature of the small *Arabidopsis* genome. The sequence of one *Arabidopsis* gene, alcohol dehydrogenase (ADH), indicates that yet another means to a smaller genome is to have fewer introns. While the maize ADH genes contain nine introns, the *Arabidopsis* gene contains only six (Chang and Meyerowitz 1986). In either case, the overall result is that *Arabidopsis* has a smaller gene than other plant species.

We have previously demonstrated that expression of a wild-type EPSPs gene at a much higher than normal level leads to glyphosate tolerance in petunia (Shah et al. 1986). We were interested in demonstrating that this could be done in an unrelated plant species with a different EPSPs gene. In order to accomplish this, the *Arabidopsis* EPSPs coding sequence, including introns, was fused to the CaMV 35 *S* promoter and reintroduced into *Arabidopsis* tissue. Northern blot analysis of *Arabidopsis* poly(A)$^+$ RNA indicated that the 35 *S* promoter would be expected to result in an approximately twenty-fold overexpression of EPSPs transcript as compared with the endogenous promoter. The results of glyphosate titration experiments demonstrated that transgenic tissue and plants were glyphosate tolerant up to normally lethal levels of the herbicide. consisting of the CaMV 35 *S* promoter and the nopaline synthase 3′ region (NOS)

An amplification of endogenous EPSPs expression does, then, provide a general means to allow plants to grow in the presence of a normally lethal level of the herbicide.

```
maqvsricngvqnpslisnlskssqrksplsvslktqqhpraypissswg 50
maqinnmaqgiqtlnpnsnfhkpqvpks....ssflvfgskklknsansm 46
lkksgmtligselrplkvmssvstaeKASEIVLQPIREISGLIKLPGSKS 100
lvlkkdsifmqkfcsfrisasvataqKPSEIVLQPIKEISGTVKLPGSKS 96
LSNRILLLAALSEGTTVVDNLLNSDDINYMLDALKRLGLNVETDSENNRA 150
LSNRILLLAALSEGTTVVDNLLSSDDIHYMLGALKTLGLHVEEDSANQRA 146
VVEGCGGIFPASIDSKSDIELYLGNAGTAMRPLTAAVTAAGGNARYVLDG 200
VVEGCGGLFPVGKESKEEIQLFLGNAGTAMRPLTAAVTVAGGNSRYVLDG 196
VPRMRERPIGDLVVGLKQLGADVECTLGTNCPPVRVNANGGLPGGKVKLS 250
VPRMRERPISDLVDGLKQLGAEVDCFLGTKCPPVRIVSKGGLPGGKVKLS 246
GSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYVEMTLKLMERFGVSV 300
GSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVEMTLKLMERFGISV 296
EHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVTVEG 350
EHSSSWDRFFVRGGQKYKSPGKAFVEGDASSASYFLAGAAVTGGTITVEG 346
CGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPPRDAFGMRHLRAIDV 400
CGTNSLQGDVKFAEVLEKMGAEVTWTENSVTVKGPPRSSSGRKHLRAIDV 396
NMNKMPDVAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLG 450
NMNKMPDVAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLG 446
ATVEEGSDYCVITPPKKVKTAEIDTYDDHRMAMAFSLAACADVPITINDS 500
ATVEEGPDYCIITPPEKLNVTDIDTYDDHRMAMAFSLAACADVPVTINDP 496
GCTRKTFPDYFQVLERITKH 520
GCTRKTFPNYFDVLQQYSKH 516
```

Fig. 6. Comparison of the amino acid sequences of the *Arabidopsis* (top) and petunia (bottom) EPSPs proteins. The sequences of the transit peptides are shown in *lower case letters* and the mature proteins in *upper case letters*

Further, the chimeric 35 *S*/EPSPs gene has been demonstrated to be a useful scoreable marker for plants which have inherited a functional T-DNA.

Acknowledgements. We would like to thank Steve Rogers and Robb Fraley for their ongoing support of this research. We would also like to thank Alan Lloyd for his advice in obtaining transgenic *Arabidopsis* plants.

References

Ausubel F, Bahnsen K, Hanson M, Mitchell A, Smith H (1980) Cell and tissue culture of haploid and diploid petunia "Mitchell". Plant Mol Biol News 1:26–32

Baker A, Schatz G (1987) Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. Proc Natl Acad Sci USA 84:3117–3121

Bennett M, Smith J, Heslop-Harrison J (1982) Nuclear DNA amounts in angiosperms. Proc R Soc Lond 216:179–199

Brown J, Feix G, Frendewey D (1986) Accurate in vitro splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract. EMBO J 5:2749–2758

Chang C, Meyerowitz E (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 83:1408–1412

della-Cioppa G, Bauer C, Klein B, Shah D, Fraley R, Kishore G (1986) Translocation of the precursor of 5-enolpyruvylshikimate-3-phosphate synthase into chloroplasts of higher plants in vitro. Proc Natl Acad Sci USA 83:6873–6877

Estelle M, Somerville C (1986) The mutants of *Arabidopsis*. Trends Genet 2:89–93

Frischauf A-M, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Gasser C, Winter J, Hironaka C, Shah D (1987) Structure expression and evolution of the 5-enolpyruvylshikimate-3-phosphate synthase genes of petunia and tomato. J Biol Chem (in press)

Kislev N, Rubenstein I (1980) Utility of ethidium bromide in the extraction from whole plants of high molecular weight maize DNA. Plant Physiol 66:1140–1143

Koornneef M, van Eden J, Hanhart C, Stam P, Braaksma F, Feenstra W (1983) Linkage map of *Arabidopsis thaliana*. J Hered 74:265–272

Leutwiler L, Hough-Evans B, Meyerowitz E (1984) The DNA of *Arabidopsis thaliana*. Mol Gen Genet 194:15–23

Lloyd A, Barnason A, Rogers S, Byrne M, Fraley R, Horsch R (1986) Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. Science 234:464–466

Meyerowitz E, Pruitt R (1986) *Arabidopsis thaliana* and plant molecular genetics. Science 229:1214–1218

Mousdale D, Coggins J (1985) Subcellular localization of the common shikimate pathway enzymes in *Pisum sativum* L. Planta 163:241–249

Pruitt R, Meyerowitz E (1986) Characterization of the genome of *Arabidopsis thaliana*. J Mol Biol 187:169–183

Rogers S, Klee H, Horsch R, Fraley R (1987) Improved vectors for plant transformation: Expression cassette vectors and new selectable markers. Methods Enzymol, in press

Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Shah D, Horsch R, Klee H, Kishore G, Winter J, Tumer N, Hironaka C, Sanders P, Gasser C, Aykent S, Siegal N, Rogers S, Fraley R (1986) Engineering herbicide tolerance in transgenic plants. Science 233:478–481

Southern E (1975) Detection of specific DNA sequences among fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Steinrucken H, Amrhein N (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. Biochem Biophys Res Commun 94:1207–1212

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119

Communicated by R.B. Goldberg

Received February 7, 1987 / July 23, 1987